Supplementary Material

Detailed Methods for the Provider Module

Formally, the probability (p) that provider i will stop recommending screening for individuals in age group g at time t, given that he or she does recommend screening at time t-1, is given by

$$p_t^{i,g} = \text{logit}^{-1}\left(S_t^{i,g} + P(f_t^{i,g} - q)\right). \quad \text{(S1)}$$

The term $S_t^{i,g}$ captures a provider's likelihood of stopping screening based on individual-level characteristics and past experience with friends, family, and patients (what we term 'social learning' in our model). This term evolves in time according to the equation:

$$S_{t+1}^{i,g} = B^{i,g} + \nu\left(S_t^{i,g} - B^{i,g}\right) + E_t^{i,g}. \quad \text{(S2)}$$

$B^{i,g}$ is an individual-level baseline propensity of provider i toward patient group g. The term $E_t^{i,g}$ captures a change in propensity due to experiences with friends, family, or patients' diagnoses with late stage breast cancer without screening at time t [33]. The parameter ν is a memory parameter that determines how slowly $S_t^{i,g}$ decays back to the baseline propensity $B^{i,g}$ (i.e., more recent experiences are more likely to be impactful).

The term $P(f_t^{i,g} - q)$ accounts for peer influence with $f_t^{i,g}$ defined as the fraction of provider i's peers who have stopped screening. In general, we assume that if more than a fraction q of provider i's peers have stopped recommending screening, this will increase the likelihood that provider i will stop screening, but if fewer than a fraction q have stopped recommending screening, peer effects will make provider i less likely to stop screening. Our assumptions here are informed by social network research on whether social influence in

different cases operates via a simple contagion or a complex contagion.[1, 2] In a simple contagion model -- what we call 'provider exposure' -- each exposure to a behavior in the network independently contributes to an individual's probability of adopting that behavior. To model provider peer influence as a simple contagion, we set $q=0$ such that the provider peer influence term becomes $P * f_t^{i,g}$, and is always positive as long as $f_t^{i,g}$ is positive. In a complex contagion, the individual is more likely to adopt a behavior once a certain threshold of peers have adopted. To model provider peer influence as a complex contagion, we set q=0.5. Under this assumption, providers are more likely to stop screening if half of peers have stopped recommending screening, and less likely to stop screening if fewer than half of peers have stopped. We label this type of peer influence 'provider norms'.

Providers are connected to each other by a random network (e.g., randomly connected to 10 percent of other providers) and each provider has a patient panel of 100 patients; we chose this panel size for computational tractability. To ensure that providers in the model learned about cancer diagnoses from patients at similar rates to those we observed in our prior provider survey, [34] we connected each provider to additional patients whose cancer diagnoses they could observe and be influenced by. We first developed a set of baseline assumptions that did not incorporate social learning ($E_t^{i,g} = 0$) or peer effects ($P = 0$ in equation 1) and calibrated our model – adjusted the parameters in the model -- to conform to rates of screening observed in our provider survey. In our base model, we randomly sampled 10 percent of providers to be "fast de-implementers"; the remaining 90 percent were "slow de-implementers".

Estimating *P*, the provider peer effect

In this section, we describe how we estimated *P*. If we re-write equation S1 as:

$$\text{logit}(p_t^{i,g}) = \left(S_t^{i,g} + P(f_t^{i,g} - q)\right), \quad \text{(S3)}$$

we see that *P* is the difference in log-odds of the probability of stopping screening between a provider who has had all of his or her peer stop screening compared to one who has had none of his or her peers stop screening patients in age group *g.* In our cross-sectional provider survey, it is not possible to determine whether associations between respondent and peer behavior are due to influence or homophily (the tendency of similar individuals to associate with each other). Therefore, to estimate the peer influence effect, we used the results from our prior study of provider peer effects magnetic resonance imaging (MRI) and positron emission tomography (PET) in breast cancer[3]. This was a longitudinal study that used SEER-Medicare data to observe changes in provider behavior over time. We used the study's estimates of the likelihood that a provider had not adopted MRI in the baseline period (2004-2006) would adopt MRI for breast cancer imaging in the follow-up period (2007-2009) based on the percent of physician peers who had adopted MRI in the baseline period. These results are shown in Table S1. We estimated the linear equation:

$$y_i = 3Pf_i + a + \varepsilon_i \quad \text{(S4)}$$

In equation S4, each *i* is an observation of a set of providers who had not adopted MRI in the baseline period grouped by the adoption rate in their peer group. These observations are listed as rows in Table S1. In the term $3Pf_{i,}$ *P* is the provider peer effect that we are trying to estimate and the 3 accounts for the fact that *P* is a log-odds of adopting per year, and there are 3 years between the baseline and follow-up period. The variable f_i is the fraction of peers who had adopted MRI in the baseline period, and is the midpoint listed in the "Baseline peer group use, proportion" column in Table S1. The dependent variable y_i is a log-odds of adopting MRI in the follow-up period. We estimated equation S3 twice – once using the log-odds upper bound from

Table S1 as the dependent variable y_i in the regression and once using the log-odds lower bound as the dependent variable. The parameter a represents the log odds of adopting over the 3-year time period in the absence of any peer effects and ε_i is an error term.

Baseline peer Group use, proportion	OR 95% CI lower bound	OR 95% CI upper bound	Log-odds 95% CI lower bound	Log-odds 95% CI upper bound
0	1 (ref)	1 (ref)	0 (ref)	0 (ref)
<=0.02 (midpoint 0.01)	0.63	2.73	-0.46	1.00
0.02-0.05 (midpoint 0.035)	0.85	2.47	-0.16	0.82
0.05-0.10 (midpoint 0.075)	1.32	3.47	0.28	1.24
>0.10 (excluded from analysis)	1.39	4.39	0.33	1.48
Regression equation			y_i=6.25f_i-0.27	y_i = 11.94f_i+0.41

Table S1 Odds ratios (OR) in this table are reproduced from the MRI section of Table 3 in (Pollack et. al., 2017).

Estimates from regression equation S3 give us estimates for *P* of 2.08 to 3.98, which we rounded to 2 to 4 for our analysis. We recognize and would expect that the different clinical scenarios—mammographic screening and breast imaging for patients to evaluate patients with known breast cancer—would likely be linked with different rates of de-adoption. However, in the absence of other evidence, this scenario was designed to give us a range of plausible values for physician peer influence.

Calibration of Baseline Assumptions

We first developed a set of baseline assumptions that did not incorporate social learning ($E_t^{i,g} = 0$) or peer influence ($P = 0$ in equation 1). We defined t=0 in our model to be 2009, the year in which USPSTF revised its breast cancer screening guidelines to recommend less screening. In our calibrated baseline model, 65% of providers recommend screening to women ages 75 and older at time t=7. This corresponds to the results of our provider survey, which we fielded in 2016 in which we found that 67% of providers recommended screening to women in this age group. We used the parameters we estimated for recommendations to older women to

model provider recommendations to younger women as well. Therefore, 65% of providers in the base model recommend screening to women ages 40-49 at t=7. In the base model, we randomly selected 10 percent of providers to stop screening younger and older women at a faster pace than their colleagues ('fast de-implementers'); this decision was motivated by our calibration process. This assumption aligns fairly well with the classic model of diffusion of innovations, in which 16 percent of the population falls into the "innovators" or "early adopters" categories, which are the first two groups to change their behavior in that model[4]. For fast de-implementers, we set $B^{i,g}$=1, which corresponded to them having an approximately 70% chance of stopping screening for younger and older patients each year in the absence of network effects. All other providers had $B^{i,g}$=-3, which meant that they had a 5% chance of stopping screening younger and older patients each year in the absence of network effects.

Calibration of the Number of Patient-Provider Connections

In BC-SAM, we use patient panels of 100 women over age 40 to make the model more computationally tractable than larger patient panels that would be more typically found in primary care, the average of which is estimated to be around 1,500[5] -2,500 patients[6]. We note that the exact number is a matter of debate; and estimates at the higher end of the range of received some criticism[7] while estimates at the lower end of the range are fairly consistent with back-of-the envelope estimates based the ratio of the population of the United States to the number of practicing primary care providers.[1]

[1] For example, the population of the United States in 2010, estimated to be 309 million, divided by the number of practicing primary care physicians, estimated to be 209,000. Facts PCW. Stats: Overview. Agency for Healthcare Research and Quality. Rockville, MD: US Department of Health & Human Services 2012., in the United States in 2010, yields 1479 individuals per primary care physician.

Our patient panels with only 100 women ages 40 and older could lead to the providers in our model having fewer patients diagnosed with cancer than providers would typically experience in the real world. To account for this, we allowed providers to be connected to, and learn about the experiences of, additional patients who they did not treat, and who they did not give screening recommendations to. We calibrated this number of connections to match the number of providers who remembered learning about a diagnosis of late stage cancer in an unscreened patient that we observed in our survey. This calibration accounted for the fact that providers might forget learning about some late stage cancer diagnoses; we used the memory parameter of $1 - \nu$, where the estimate for ν based on our prior work[9]. As a result of this calibration process, we connected each provider to 325 patients they do not treat in addition to the 100 patients they do treat. With this number of additional connections, 11% of providers in the BC-SAM remembered learning about a late-stage diagnosis made in an unscreened patient seven years after de-implementation began compared to the 10% of providers who reported knowing such patients in our CansNET survey.

Calibration of the Patient, Friends and Family Network Effect

The term $E_t^{i,g}$ captures a change in propensity due to experiences with friends, family, or patients' breast cancer diagnoses at time *t.* We calibrate this term based on our survey results which found that providers who reported friends, family, or patients diagnosed with late stage disease without screening were significantly more likely to recommend screening. In our survey, we did not find that this association varied significantly based on the type of relationship (friend, family, or patient) so grouped all three together in our model We found that $E_t^{i,g} = -3$ gave us good agreement with the results from BC-SAM. Using this value, in the model we found that providers who remembered having a patient, friend, or family member with late stage diagnosis, not diagnosed through screening were more likely to continue to recommend screening at 7 years after the start of de-implementation with an odds ratio of 2.3. This value gives us good agreement with the results from our CanSNET survey, in which we found that learning about a social network member diagnosed with late stage breast cancer who had not been screened increased the likelihood of recommending breast cancer screening to both younger (OR 2.14, 95%CI 1.07, 4.26) and older (OR 2.44, 95%CI 1.40, 4.25) patients.

Validation of the Number of Patient-Provider Connections

The calibration described in the previous section found that providers should have about 425 women ages 40 and older in their patient panel in order for providers to recall learning about late-stage diagnoses of breast cancer made without screening at rates similar to those we found in our CansNET survey. Approximately 25 percent of the total U.S. population is women ages 40 and older[10], which we can use to calculate an implied average total patient panel size of 425/0.25 = 1,700, which is consistent with the estimated average patient panel range for primary care providers of 1,500-2,500 that we described above. Therefore, we find that we estimate a

reasonable average patient panel size using from (1) the frequencies with which patients who have not been screened are diagnosed with late-stage cancer within BC-SAM, (2) rates at which providers learn about such outcomes, based our provider survey, and (3) differences in provider recommendations among providers who do and do not recall learning about such an outcome in a patient and (4) a memory parameter we estimated based on the results of our prior patient survey.

Fast and slow de-implementers

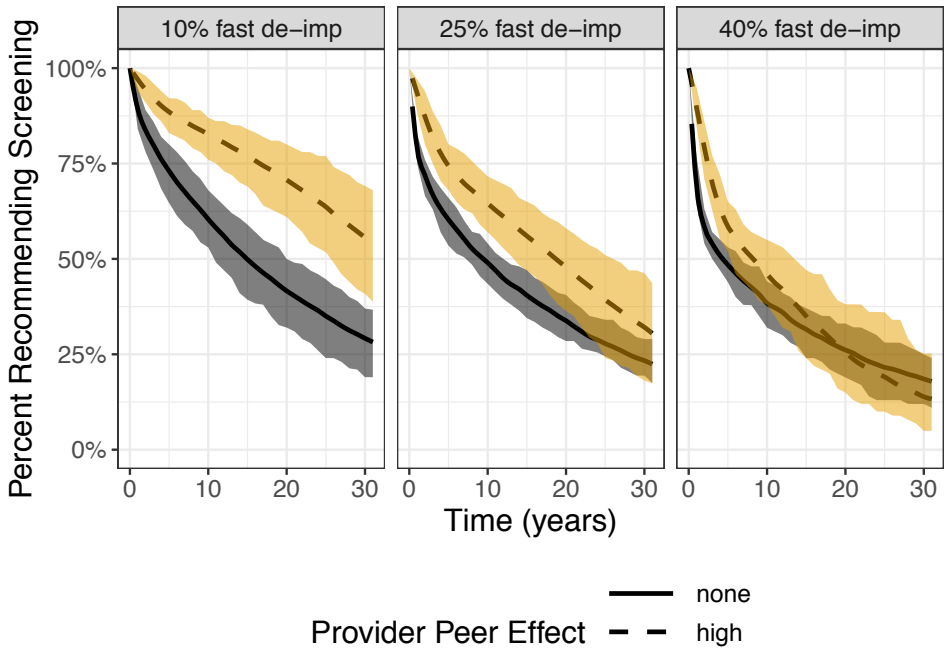
Our models include two groups of physicians: fast- and slow de-implementers. Most diffusion of innovation models include more heterogeneity with 4 or 5 groups (parsing people into categories such as innovators, early adopters, early majority, etc.).[11] However, we opted for the simpler model as adding more groups would have increased the complexity and number of parameters in the model without improving accuracy as we do not have a way to estimate these additional parameters from the data we previously collected.

Sensitivity analyses

In sensitivity analyses, we varied key parameters and decisions to show their impact on our overall model. Our main model makes assumptions on the proportion and rate of fast de-implementers as well as the de-implementation rate of slow de-implementers that were consistent with the results of our CansNET survey. We recognize that there is significant uncertainty around our choices of parameters because CansNET gives us provider recommendations rates at only a single point in time, and, without good national rates of screening in 2009, we assumed that all providers recommended screening to older and younger women before 2009. Therefore, we (1) changed the proportion of providers who we classified as fast and slow de-implementers, (2) altered the speed at which fast de-implementers give up screening, and (3) varied the baseline

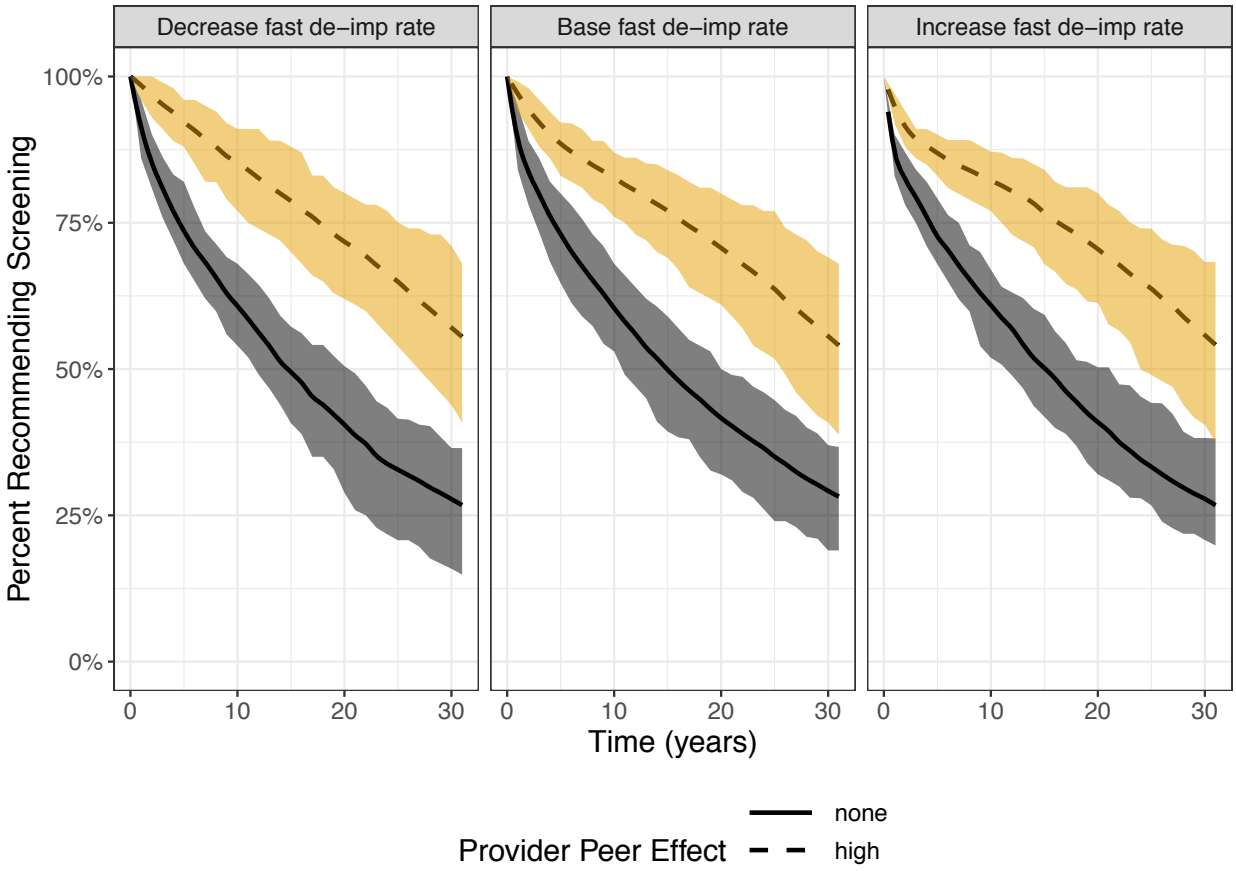
de-implementation rate of slow providers. For each sensitivity analysis, we run the models with and without incorporating provider peer effects, which we model as peer norms (e.g. where providers are influenced to copy the behavior of the majority of their peers).

The analyses suggest that increasing the proportion of providers classified as fast de-implementers reduced breast cancer screening over the 30-year time horizon (Figure S1). Changing the changing the speed with which fast de-implementers stopped recommending screening (while keeping the proportion of fast de-implementers fixed) had little impact on the de-implementation dynamics in the scenarios we consider (Figure S2). In contrast, changing the baseline de-implementation propensity for slow de-implementers to stop recommending screening had a substantial effect on our estimates (Figure S3). Each is described in greater detail below.



- *Figure S1: Impact of the percentage of providers who are fast de-implementers ("fast de-imp"). Curves with a solid line have no provider peer effects; the dashed line shows a high provider peer effect (P=4). In these scenarios, $B^{i,g} = 1$ for fast de-implementing providers and -3 for slow de-implementing providers. These scenarios include the influence of patient experiences on provider decision making.*

Figure S1 shows how provider peer effects impact provider recommendation rates under different assumptions about the percentage of providers who are fast de-implementers at baseline. Across the three panels, we observe that increasing the proportion of fast de-implementers leads to lower rates of recommendations over time, with steeper initial declines. The effects are more pronounced in models that do not incorporate provider peer effects (with the exception of the third panel where the proportion of fast de-implementers is the highest and, towards the end of the time period where nearly all providers no longer recommend screening). This is likely due to provider peer effects working to temper de-adoption by reinforcing existing norms.

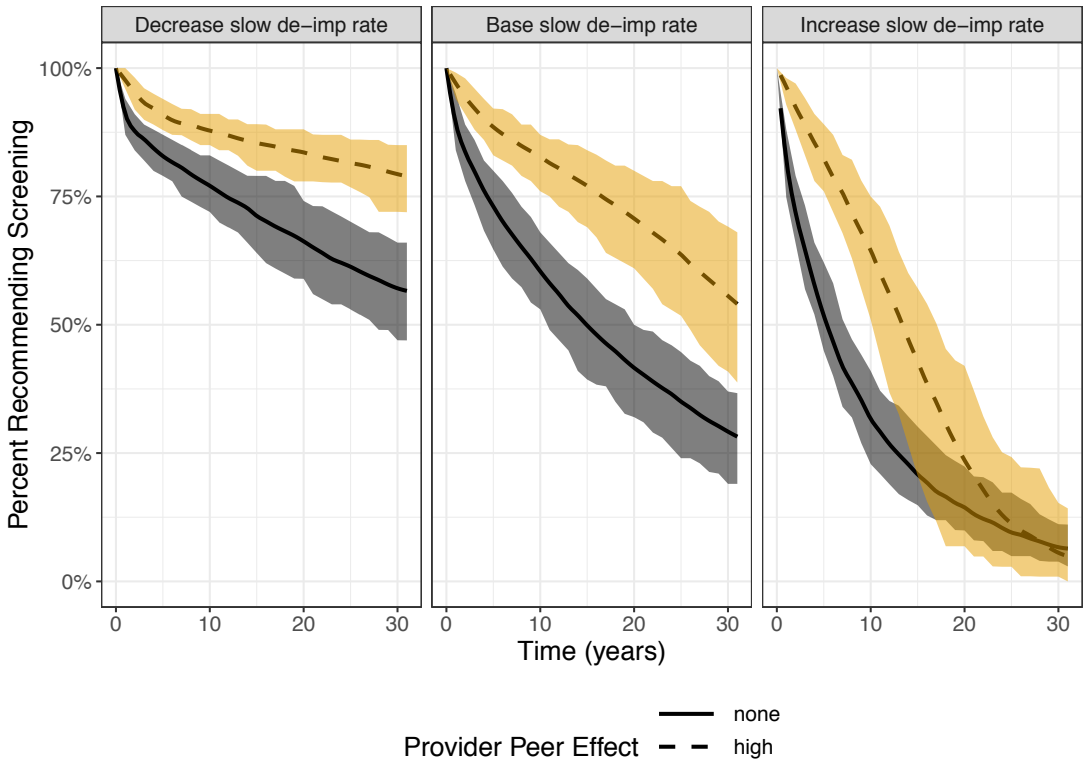


- *Figure S2: Impact of the baseline rate $B^{i,g}$ at which fast de-implementers stop screening on recommendation rates. In our "decrease fast de-imp" scenario, fast-de-implementers have a baseline de-implementation propensity of B=0, which corresponds to a 50% annual probability of stopping screening in the absence of network effects The base fast de-implementation rate has B=1 for the fast-de-implementers, which is an annual de-implementation rate of 73% in the absence of network effects, and the "increase fast de-imp" scenario has B=2 for fast de-implementers, which corresponds to an annual de-implementation rate of 88% in the absence of network effects. Curves with a solid line have no provider peer effects; the dashed line shows a high provider peer effect (P=4). In these scenarios, $B^{i,g}$ = -3 for slow de-implementing providers. These scenarios include the influence of patient experiences on provider decision making.*

-

Figure S2 shows that the speed with which fast de-implementers stop recommending screening has little impact on the de-implementation dynamics in the scenarios we consider. In our base case assumptions, we set our baseline de-implementation propensity for fast de-implementers to *B=1.* In sensitivity analysis, we increase this to *B=2* or decrease it to *B=0.* When *B=2,* about 88% of fast de-implementers stop recommending screening after one year in the absence of network effects. When *B=0,* it takes 3 years for 88% of fast de-

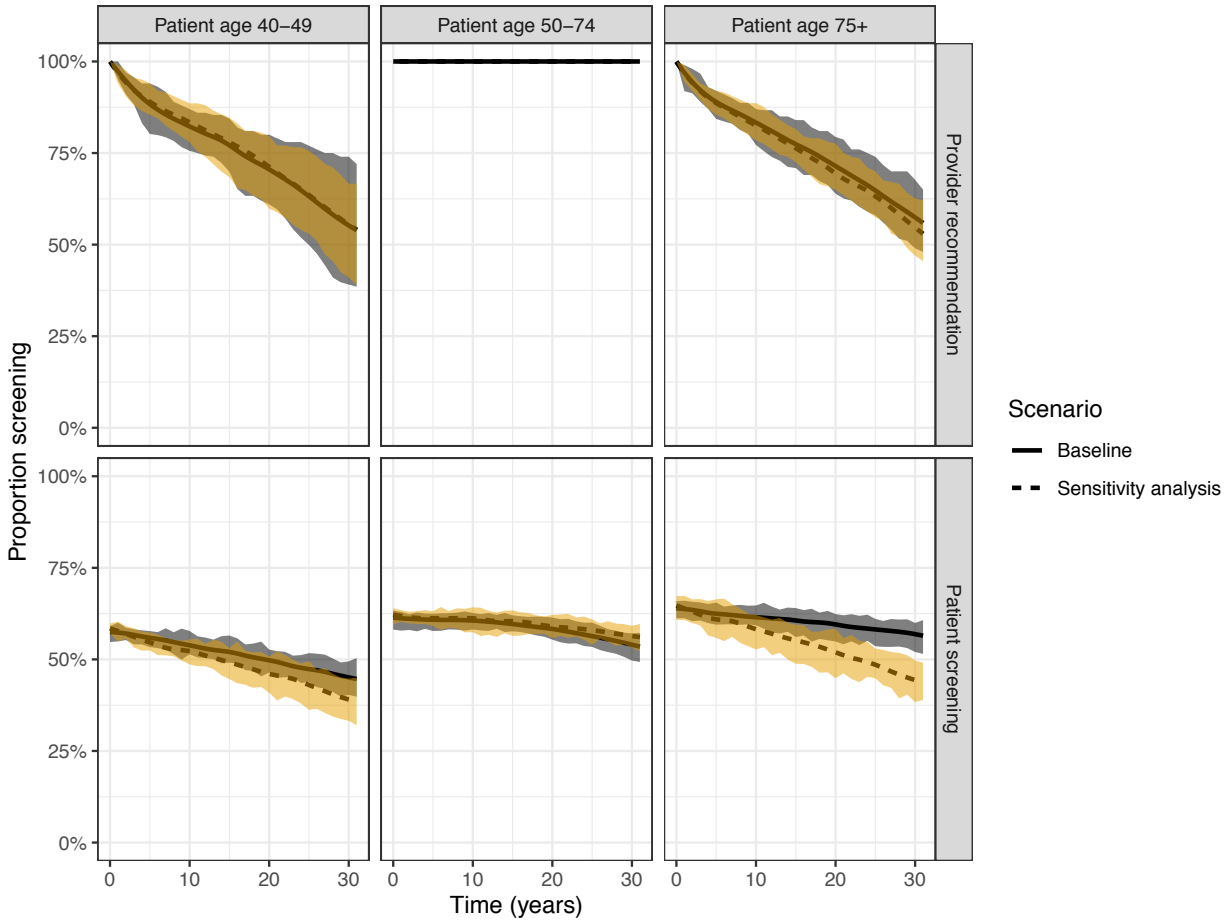
implementers to stop recommending screening in the absence of network effects. Because we are interested in dynamics over a 30-year time period, and slow de-implementers stop recommending screening at an annual rate of about 5% in our base case, our overall dynamics do not change much if it takes 1 year or 3 years for most fast de-implementers to stop recommending screening. Across all the panels, we note that incorporating provider peer effects (modeled as peer norms) leads to higher rates of recommending screening, again due to these norms ability to reinforce existing behavioral patterns.



- *Figure S3: Impact of the baseline rate $B^{i,g}$ at which slow de-implementers stop screening on recommendation rates. In our "decrease slow de-imp" scenario, slow de-implementers have a baseline de-implementation propensity of B=-4, which corresponds to a 2% annual probability of stopping screening in the absence of network effects. The base slow de-implementation is $B^{i,g}$=-3 for the slow de-implementers, which is an annual de-implementation rate of 5% in the absence of network effects, and the "increase slow de-imp" scenario has $B^{i,g}$=-2 for fast de-implementers, which corresponds to an annual de-implementation rate of 12% in the absence of network effects. Curves with a solid line have no provider peer effects; the dashed line shows a high provider peer effect (P=4). In these scenarios, $B^{i,g}$ = 1 for*

fast de-implementing providers. These scenarios include the influence of patient experiences on provider decision making.

Figure S3 shows that the baseline de-implementation propensity for slow de-implementers stop recommending screening has a substantial impact on de-implementation dynamics over a 30-year period in BC-SAM. In our base case assumptions, we set our baseline de-implementation propensity for slow de-implementers to $B^{i,g}$=*-3.* In sensitivity analysis, we increase this to $B^{i,g}$=*-2* or decrease it to $B^{i,g}$=*-4.* When $B^{i,g}$=*-2,* about 12% of slow de-implementers stop recommending screening after one year in the absence of network effects. When $B^{i,g}$=*-4,* it takes about 7 years for 12% of slow de-implementers to stop recommending screening in the absence of network effects. As with the prior models, peer effects work to decrease the speed of de-implementation with the exception of the third model at the end of the time period when overall rates of recommendations are quite low.



- *Figure S4: Impact of the provider recommendation effect. The solid line shows the baseline provider recommendation effect, and the dashed line shows the sensitivity analysis in which the provider recommendation effect is doubled.*

Figure S4 shows a sensitivity analysis in which we double the size of the provider recommendation parameter. At baseline, based on our prior analysis, the increase in log-odds for a patient with a provider recommendation to be screened is 0.89. In the sensitivity analysis, we increased the impact of the provider recommendation to an increase in log-odds of 1.78. Compared to baseline, more women ages 50-74 are screened in the sensitivity analysis and fewer women ages <50 or 74+ are screened. In other words, women's screening patterns are more responsive to changes in provider recommendations in the sensitivity analysis presented here.

References

1. Monsted B, Sapiezynski P, Ferrara E*, et al.* Evidence of complex contagion of information in social media: An experiment using Twitter bots. Plos One 2017;12(9).

2. Centola D, Macy M. Complex contagions and the weakness of long ties. American Journal of Sociology 2007;113(3):702-734.

3. Pollack CE, Soulos PR, Herrin J*, et al.* The impact of social contagion on physician adoption of advanced imaging tests in breast cancer. JNCI: Journal of the National Cancer Institute 2017;109(8):djw330.

4. Rogers EM. *Diffusion of Innovations*. 4th ed. New York: Free Press; 1995.

5. Weiner JP. Prepaid Group Practice Staffing And US Physician Supply: Lessons For Workforce Policy: What can we learn from examining the staffing levels of some of the country's largest organized delivery systems? Health Affairs 2004;23(Suppl1):W4-43-W4-59.

6. Murray M, Tantau C. Same-day appointments: exploding the access paradigm. Family practice management 2000;7(8):45.

7. Raffoul M, Moore M, Kamerow D*, et al.* A primary care panel size of 2500 is neither accurate nor reasonable. The Journal of the American Board of Family Medicine 2016;29(4):496-499.

8. Facts PCW. Stats: Overview. Agency for Healthcare Research and Quality. Rockville, MD: US Department of Health & Human Services 2012.

9. Nowak SA, Parker AM. Social Network Effects of Nonlifesaving Early-Stage Breast Cancer Detection on Mammography Rates. American journal of public health 2014;104(12):2439-2444.

10. Bureau UC. *U.S. and World Population Clock*. https://www.census.gov/popclock/.

11. Valente TW. *Network models of the diffusion of innovations*: Hampton Press Cresskill, NJ; 1995.